DCD

Sponsored by

VERTIV

Edge Supplement

Fresh thinking for a new frontier

INSIDE

The Edge in health

> Don't prescribe every sector with the same course of treatment

Right-sizing the Edge

> Maybe racks of kit aren't the best way to support remote applications

Edge of Mars

> Seeking signs of life on our neighbor, Perseverance is the most extreme Edge project ever

Because Downtime Is Not an Option.

No Matter Where Your Edge Is.

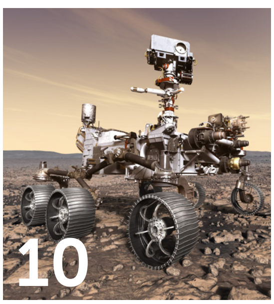
When your network edge is hard to get to, give your team access and secure IT management.

What's Your Edge?

Vertiv.com/DCDEdge

Sponsored by

Contents



Honing a fresh Edge

Each time we return to the subject of Edge, it becomes ever more clear that there is still plenty of thinking to be done about it.

We've been told many times that Edge is a frontier territory, where infrastructure has to support localized applications that need responsive, low-latency processing.

So it shouldn't be a surprise when the old techniques don't always match the new demand. This supplement looks at some ways to understand Edge, and make it work, on this planet and beyond.

Decoding the Edge

Edge includes a variety of use cases and applications or "archetypes" within the blanket term "Edge."

It's important to understand and respond to the requirements of each specific situation (p6).

The healthcare Edge

As the past year has sadly shown, nothing is more important than healthcare - but for the industry to advance, it will need more compute, at lower and lower latencies.

Edge deployments at hospitals and pharmaceutical companies are on the rise, but one size does not fit al (p4).

The distributed Edge

Edge applications are not quite the same as cloud applications. They don't need exactly the same hardware.

The network may take the lead; small identical nodes on a mesh might do just as good a job as traditional racks, and they might do it more resiliently.

The open, networked Edge

Taking that thought further, what kind of network do we need?

5G is a convenient catch-all phrase for the next generation mobile networks, which will surely be there when we interconnect our Edge applications.

But what kind of 5G will we have? There's a race on - to create an open, standards-based set of protocols called OpenRAN, before the urgent need for bandwidth forces everyone to install the network vendors' latest proprietary offerings, and usher in another era of high-cost infrastructure.

OpenRAN could win the race, and we might all benefit (p13).

The extra-terrestrial Edge

Finally, as a reminder of just how revolutionary Edge technology is, we take you to another world.

NASA's Perseverance rover is exploring Mars. It is the most remote application imaginable, with 20-minute latency times that would cripple any traditional remote control systems (p10).

Hands-on repairs and real-time control have been literally impossible, during a risky one-off landing and a lengthy search for signs of life on our nearest neighbor in the solar system.

Do these extreme conditions need cutting edge tech? Nope. The rover is powered by a processor out of a 20-year old Macintosh. Its drone copter carries a processor from a six year old smartphone, which your child would scorn.

As we said at the start, Edge needs fresh thinking.

On Mars, that led to a choice of trusted technology and appropriate tools.

Edge computing in the health sector

Nick Booth
Contributor

Don't prescribe every sector with the same course of treatment

The health computing sector is suffering high data deposits and networking sclerosis. A prescription of Edge computing might help the patient pull through, but the treatment will need experts to administer it without serious side effects.

Unfortunately, the health sector is experiencing rapid infrastructure metamorphosis, exacerbated by Covid-19, which could severely impede the thought processes of any organization.

It's a widespread problem, which has been diagnosed in many industry sectors, for instance by Digital Realty's Data Gravity Index, which reveals that large amounts of data, stored without due care, can create stasis and resistance to change in organizations.

The patient (global computing in all sectors) has a data metabolism of 1.5 exaflops but will need to show massive levels of improvement if it can summon the extra 9 exaflops needed to process the 15 zettabytes of information that will be stored in its brain by 2024, according to Interxion's analysis.

In short, without support, the extremities of the network may experience the IT equivalent of peripheral neuropathy.

The syndrome is particularly acute in the health sector because of two of the many reactions to Covid-19: remote working and research.

Permanently remote

The trend for remote consultancy won't be reversed post Covid-19. A permanent consequence of lockdown is that installations to support digital therapeutics will increase by 69 percent every year, according to Juniper Research, which projects revenues of $53.4 billion by 2025. Estimates of the projected volumes of medical research data fluctuate wildly because growth is so furious with the likes of GlaxoSmithKline building 400 petaflops supercomputers across Europe to speed drug research.

Suffice it to say that bodies such as the International Covid-19 Data Alliance (ICODA) see IT infrastructure problems, rather than data volumes, as the primary challenge to its progress, according to an ICODA seminar given in March 2021.

Edge computing cannot be a universal panacea for the health sector because this is a category of service with a complicated history. And the challenges vary according to culture and infrastructure.

In the UK's National Health Service (NHS), the management culture can be harder to reconcile than power or cooling or comms connections.

While private enterprises will have an ops manager and a facilities manager with well-defined roles, a hospital has IT staff who know nothing about power, cooling and comms racks. This is not a problem, if the service provider can manage these decisions for the client, but getting the facilities laid on is a massive initial challenge.

Get a room

Just finding a place in a hospital for a small self-contained micro data center is a problem, according to one service provider we spoke to, who has worked on installing Edge IT into hospitals: "The hardest part of our Edge computing work for the NHS is finding a room."

The biggest challenge is the time scale. An IT project leader may make an instant decision on the need for servers but suppliers can spend three months just trying to get a meeting with the right head of department. It could be the operations boss, it could be the estate manager. There is no one person cracking the whip.

In one project a hospital appointed an installer to create a local data center. The service provider advised the installer on the best options and awaited instructions while the reports were processed. At the initial consultation, the installer saw a room that was ideal for hosting a local data center. It was months later when they heard another room was allocated. Now they had to get the requisite power connected, and that took eight months. When it was all ready, the news came that the building was about to be torn down.

"At one hospital we had to walk around for ages to see if we could shoe horn something in," said an installer who asked not to be named. "If nobody gives you a space, you have to go in the car park or the corner of a room."

Monitoring and cooling

There is a general pattern of installation in hospitals: one or two big data centers (a computing site and a disaster recovery facility) with multiple discrete and informal comms rooms tucked away, each with some sort of uninterruptible power supply (UPS). There can be 200 UPS systems dotted around the hospital campus and these are hard to find and impossible to manage.

The upside is that this creates more sales for service providers; they can monitor them remotely via the inbuilt DCIM platform and bring all the information together for the client.

Providing containerized data centers that use liquid cooling, as part of an Edge computing package, solves multiple problems in the case of a hospital. Since it immerses the servers in liquid, that prevents any likelihood of dust fires ignited by overheated circuitry (dust is a major problem in the type of rooms that will be on offer in a hospital). Immersion also stops any flakes of metal from being blown onto the boards and short circuiting the computer.

"At one hospital we had to walk around for ages to see if we could shoe horn something in. If nobody gives you a space, you have to go in the car park."

AI for care

In private health service providers, where the decision making is quicker, the challenge is to be sufficiently adaptable to circumstances. Care homes exemplify another aspect of how Edge Computing can solve congestion and compliance problems.

Covid-19 put a huge strain on staff in care homes, making it harder for them to find the time to keep tabs on their patients. Meanwhile, they also had to contend with new privacy laws that limited the amount of camera surveillance that could be carried out.

One solution to the problem was created at The University of Amsterdam (UoA), where Dr. Harro Stokman invented a way to use artificial intelligence to make sense of the patterns of events in each room.

The legislation restricted the time that humans can watch patients through video cameras. However, there are no such limitations on a computer and, if it is deemed intelligent enough, the machine's judgment can be trusted on the well-being of a patient. This was the logic of Stokman's Kepler Night Nurse (KNN) AI system which observes patients and decides if events (such as a fall) need intervention.

The problem is that the KNN system creates too much data to load into the cloud without creating huge bottlenecks and comms bills. In response, the UoA spin off company, Kepler Vision Technologies (KVT) built an Edge Box to handle all the data locally, using Nvidia's small form factor Jetson Xavier NX module. The Edge computing node can process data locally and improve on the quality of intelligence gathered. By localizing the analysis less data is sent to the cloud to be processed.

Infrastructure still needed

Creating Edge computing hardware is one thing. But where will the supporting infrastructure come from?

Mobile telco industry watcher Dean Bubley, founder of Disruptive Analysis, warns that the expectations created for 5G are unrealistic, especially in regard to supporting systems that need instant response times.

"The low-latency 5G Emperor is almost naked," says Bubley. In some cases, he concedes, the ultra-reliable low-latency (URLLC) associated with 5G could minimize network round-trip times for new apps and devices that need instant responses. "In that respect mobile Edge computing can cater to them, in the form of regional computing facilities or servers at each base station," says Bubley.

However, there are many new applications where the latency has to be a lot better. An endoscope or microsurgery tool might need to respond to controls and send haptic feedback 100 times a second, i.e. every 10ms. Drones are being proposed for drug transport between hospitals, but these flying devices must react in two milliseconds to a control signal, or a locally-recognized risk. It's also doubtful if 5G could offer the latency needed by photon sensors used in research, which need to operate at picosecond durations.

Fiber to the rescue?

One of the US's answers to the infrastructure challenge is open access or competitive fiber optic network, such as SiFi Networks' FiberCity offering. SiFi promises this allows access to multiple service providers and geographically diverse paths on a fiber optic network that gives '99.9999' percent reliability.

In this model, a city-wide fiber network passes by each home and business and effectively gives each company a private network. This could create a citywide private network, secure from the Internet, to deliver data to hospitals and research bodies via high-speed symmetrical connectivity.

This would give them much less of a problem sending data over the cloud, according to SiFi Networks CEO Ben Bawtree-Jobson.

Storing files locally creates issues for sharing with consultants offsite and internationally. Suitably sized cloud storage could come closer to seamless collaboration between consultants.

"The question is then about how much bandwidth is required and 100 percent fiber optic networks solve this problem," says Bawtree-Jobson.

Meanwhile, Juniper Research reports that mobile players are partnering across the globe to build the mobile Edge computing infrastructure. Between them the likes of AT&T in the US, LG/Google in South Korea and the 5G Future Forum are spending $8.3 billion, by 2025, on the networking equivalent of life support systems for all those Edge systems.

So the patient's prospects are looking better. But keep an eye on the monitors.

Bringing the Edge down to size

Edge applications aren't the same as cloud ones, so they don't need the same hardware

Dan Robinson
Correspondent

Edge computing has been one of the major trends of the past several years, as applications have started to require lower latencies, and the volume of data handled by endpoint systems has grown to the point where streaming it all back to a cloud data center may be too costly, slow and bandwidth-hungry.

But one of the issues with Edge computing is that it is a fairly nebulous term that means different things to different people. Does the edge of the network refer to endpoint devices, or to the communications equipment that links such devices back to the core, or does it cover both of these examples and more?

Gartner, for example, defines Edge computing as solutions that facilitate data processing at or near the source of data generation, but goes on to add that Edge computing serves as the decentralized extension of the campus networks, cellular networks, data center networks or the cloud.

For the telecoms industry, edge computing has been closely identified with the development and deployment of 5G networks, with their goals of handling data rates of gigabits per second, minimal latency, and the ability to support a large number of simultaneously connected endpoint devices. These requirements are expected to see cellular base stations increase their amount of compute power so that they effectively become miniature data centers.

Meanwhile, enterprises and service providers have also been investing in so-called micro data centers in order to serve the needs of Edge computing. These micro data centers vary in size, but a typical product

> *Micro data centers are good for factories, but not for every use case*

is the equivalent of a data center rack with power distribution units and cooling encased in a protective enclosure, which can be populated with standard rack-mount servers, storage and switch kit.

Such solutions are perfect in a factory setting, for example, where a significant amount of compute power is required to monitor and control production lines, especially where multiple machine vision systems are employed, and fixed wiring is likely to be already in place for communications and power.

However, Edge computing covers such a broad range of applications and use cases that no one solution fits every problem, so a broad spectrum of capabilities is needed to fit every niche, and many will need to be more compact and have different capabilities.

"There's actually a hierarchy of processing that you would want as you move from the edge of the network all the way into the core," says Kurt Michel, senior vice president of marketing for Edge infrastructure firm Veea.

Veea develops what it refers to as smart edge nodes, which can start with a deployment of just a single node but can scale by adding more nodes if required, as nodes can communicate with each other via a built-in mesh networking capability. Each node is a tiny box that looks like a WiFi access point, but contains a 64-bit quad-core Arm processor running Linux.

According to Michel, this model emphasizes both computing and connectivity, which is important for Edge applications, but the nodes can operate as if they were a single system via mesh networking.

"These separate nodes, you deploy them, and they will connect to each other. And what they do is they basically create a single virtual, connected compute platform. And they can connect to all of your different IoT-type devices, so cameras, thermal sensors, air quality sensors, vibration sensors, and the ways they connect might be Bluetooth, or LoRaWAN or ZigBee, or WiFi, or just plain old physical Ethernet," he says.

Because the hubs operate as a distributed system, any IoT devices connected to any of the nodes is visible to and can be accessed by applications running on any of the other nodes. It also means that the devices can share workloads.

"The applications themselves run in Docker containers. And that makes these applications incredibly portable. So you can move them from one node to another node. And if you find a particular node becoming overwhelmed, you can deploy another node in that location," Michel explains.

One upshot of all this is that a mesh network can provide a decent amount of aggregate processing power if needed - perhaps as much power as a micro data center - but that is not the way they are intended to be used. Instead, they are aimed at fitting into locations such as smart buildings, retail outlets or outdoor smart city environments, in sites where there may not be the space or power available to support a micro data center.

The range of applications that such devices might be used for is diverse. Michel cites the example of a retail outlet that might have a node connected to a security camera monitoring the entrance to the premises. The device could run a machine learning visual recognition model to detect people entering and whether they are wearing a Covid face mask, and generate an alert if not.

> *"You have just got to find the balance, you basically take your tasks, and you break them up into the things that need a rapid response and the things that require deeper processing."*

This hypothetical example illustrates some of the justifications for such edge deployments; streaming the video back to a cloud data center for processing may introduce unnecessary delays in generating a response, and incur unnecessary costs in network bandwidth.

"Anything that requires real time responsiveness, any control systems for robotic systems, industrial factory settings, whatever, all that real stuff that really can't handle the delay that going back to the cloud gives you," Michel says. "You have just got to find the balance, you basically take your tasks, and you break them up into the things that need a rapid response and the things that require deeper processing."

It isn't just specialist vendors that are looking to address the broad spectrum of device requirements that edge deployments encompass. In March, Lenovo expanded its range of ThinkEdge systems with a pair of ruggedized devices, the ThinkEdge SE30 and ThinkEdge SE50. Both are essentially PC hardware in compact enclosures designed for harsh industrial environments, but can be configured with 4G or 5G wireless modules in addition to WiFi, and feature RS232/422/485 serial ports for industrial peripherals.

However, products such as these largely leave it up to the user or a systems integrator to provide a suitable software stack for their Edge computing application, whereas a specialist like Veea offers a turnkey Edge node platform that allows the user to focus on making their application work.

Edge computing has been enabled by advances in computing that make it possible to add intelligence almost anywhere, and also by the spread of pervasive communications networks. But organizations need to take care when deciding whether Edge or cloud is the best place for data processing to happen, and also when choosing an appropriate platform from the wide choice available.

Decoding the Edge: A Systematic Approach to Edge Deployments

By Alex Pope, Vice President, Integrated Rack Solutions – EMEA, Vertiv

According to Ericsson, global mobile data traffic is estimated to reach 226 exabytes per month sometime in 2026. Let's put that in context. An exabyte is 1 billion gigabytes. If you collected and stored all the words spoken in human history, that would equal about five exabytes. If you did so 45 times over, you would have 226 exabytes – the amount of data we'll be generating each month just five years from now.

The applications driving this growth range, from streaming videos and gaming, to telehealth and pandemic-driven remote work, to pilot projects for autonomous vehicles. The disparate technologies making it all possible are linked in one critical way: their increasing reliance on computing at the edge of the network.

The edge of the network presents a number of unique challenges. Whereas traditional data centers are somewhat homogeneous – different in size and details, but unquestionably data centers – the edge is comprised of a universe of small IT spaces, ranging from the single-server IT closet to far more sophisticated cloud deployments. As these sites have become more and more critical, they have become more complex, and today's edge bears little resemblance to the earliest distributed sites.

One of Vertiv's first steps toward bringing some order to this new world was to categorize edge sites based on applications they support. We started by examining dozens of edge use cases, focusing on workload requirements and corresponding needs for performance, availability and security. Ultimately, we identified four edge archetypes. We use these models to better understand and equip edge sites to meet the needs of the organizations and end users that rely on them. The four archetypes are:

- **Data Intensive:** This includes use cases where the amount of data makes it impractical to transfer over the network directly to the cloud or from the cloud to point-of-use due to data volume, cost, or bandwidth issues. Examples include smart cities, smart factories, smart homes/buildings, high-definition content distribution, high-performance computing, restricted connectivity, virtual reality, and oil and gas digitization.
- **Human-Latency Sensitive:** This archetype includes use cases where services are optimized for human consumption, and it is all about speed. Delayed data delivery negatively impacts a user's technology experience, potentially reducing a retailer's sales and profitability. Use cases include smart retail, augmented reality, website optimization, and natural language processing. Increasingly, these applications are becoming the way people interact with

brands, institutions, and each other.

- **Machine-to-Machine Latency Sensitive:** Speed also is the defining characteristic of this archetype, which includes the arbitrage market, smart grid, smart security, real-time analytics, low-latency content distribution, and defense force simulation. Because machines are able to process data much faster than humans, the consequences for slow delivery are high. For example, the continuous optimization of our energy consumption, quality, and usage of renewables requires speed of analytics and decision implementation on a scale only machines can achieve.
- **Life Critical:** This archetype encompasses use cases that directly impact human health and safety. Consequently, low latency and reliability are vital. Use cases include smart transportation, digital health, connected/autonomous cars, autonomous robots, and drones. For example, as transportation becomes increasingly automated, the onboard processing of the vehicles and drones will be augmented by connectivity to real-time traffic, safety, scheduling, and routing intelligence processed remotely.

Of course, applications are only one variable, and they are virtual. The physical assets enabling these applications have to actually live somewhere — locations at the edge — which tend to follow one of these four deployment patterns:

- **Geographically Disperse:** These sites are similarly sized and spread across large geographies — typically a country or region. Retail, with stores scattered across a chain's footprint, or consumer finance, which includes bank branches, are good examples.
- **Hub and Spoke:** This also typically covers a large area, such as a country or region, but the sites are organized with multiple smaller deployments around a larger hub. Communications and logistics networks tend to embrace this model.
- **Locally Concentrated:** These are smaller networks, often servicing campus settings, such as those common to healthcare, education, and industrial sites. They also tend to feature a number of small deployments connected to a larger central facility.
- **Self-Sustained Frontier:** This pattern, with widely spread footprints ranging from regional to global, consists of the largest individual edge sites. They often carry many traditional data center characteristics but tend to be of modular construction. These sites are often employed by cloud providers to serve sizable areas. Smaller versions are commonly used for disaster recovery as well.

After categorizing first by archetype, which is focused on the virtual application, then by geography, we go even deeper, slicing these populations by physical environment and corresponding characteristics of sites within a given group. This provides the final layer of site analysis and allows us to quickly and easily configure these Edge sites to meet the specific needs of our customers. The categories are:

- **Conditioned and Controlled** (<6 kW per rack or >6 kW per rack): These are purpose-built spaces that are climate controlled and secure. The only difference in sites is rack density.
- **Commercial and Office:** These are occupied spaces with existing, but limited, climate control and sites that are typically less secure.
- **Harsh and Rugged:** These require more robust systems and enclosures to protect against large amounts of particulate in the air. These often are industrial sites with the threat of water exposure and in proximity to heavy traffic or machinery. They lack climate control and are far less secure.
- **Outdoor Standalone:** These are outside and unmanned sites, exposed to the elements and requiring a shelter or enclosure. They can be in remote locations that require some time to reach for planned or unplanned service.
- **Specialty:** These sites likely share characteristics with one of the above categories but must be handled differently due to special regulatory requirements that could be tied to application, location, or other factors.

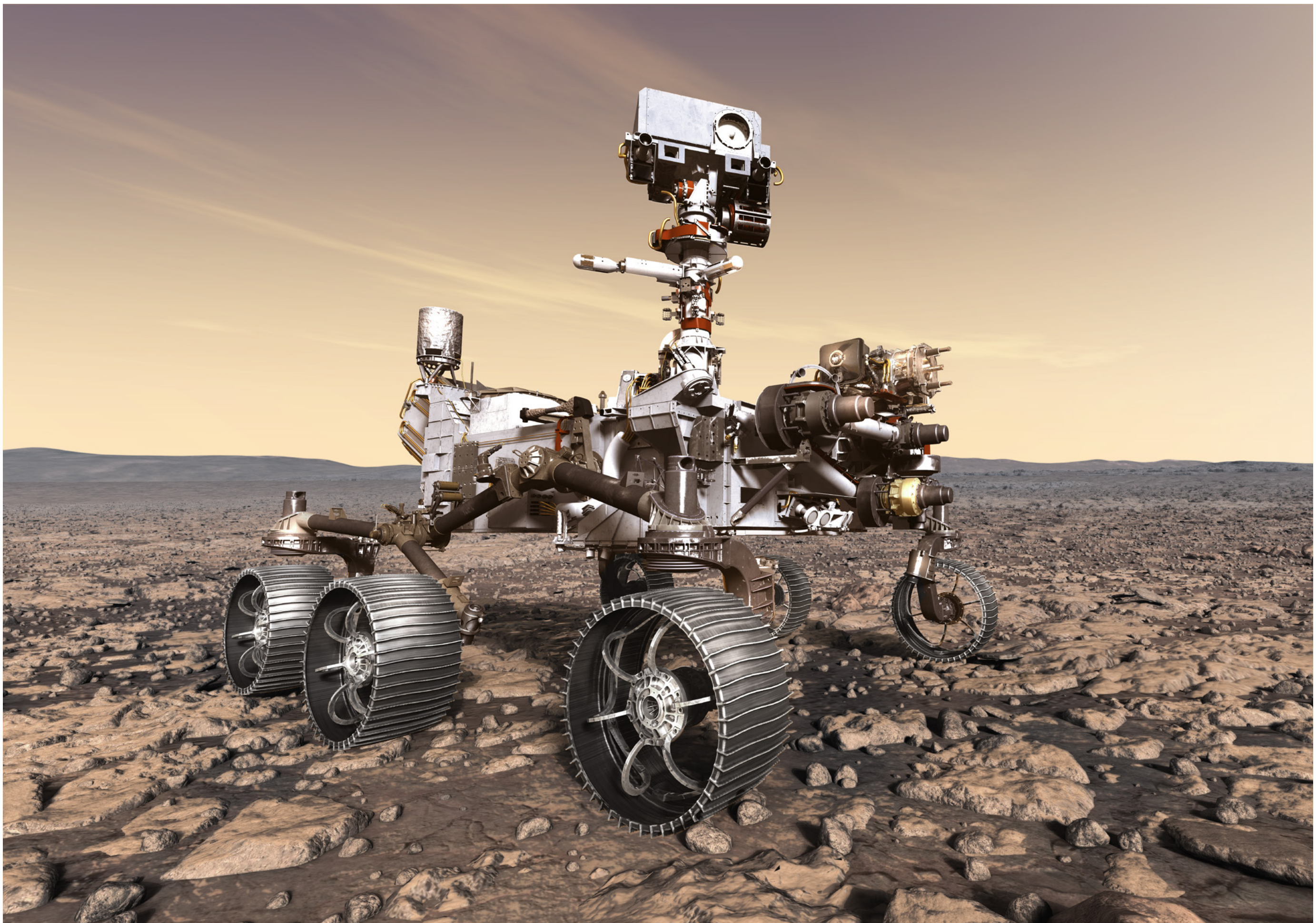
To be clear, this is not an academic exercise, but a practical methodology to understand (1) the IT functionality and characteristics each site must support; (2) the physical footprint of the Edge network; and (3) the infrastructure attributes required of each deployment. Once we have those data points, we can configure, build, and deploy exactly what is needed. We can do it faster and more efficiently while minimizing time on site for installation and service.

We have been conditioned to see the Edge as some sort of IT wilderness that can't be defined or aligned with our traditional approaches to the data center. This is not the case. By applying a systematic approach to site analysis, we can decode the Edge and take a major step toward standardizing the Edge deployment process. Ultimately, this will help our customers achieve their primary goal of reducing the time and cost required to deliver the application experience they designed for their users.

VERTIV™

Alex is grateful to have had opportunities in Vertiv to design, assemble, and lead multiple businesses in the Americas and in Europe, the Middle East and Africa. Over 13 years with Vertiv, in roles spanning marketing, strategy, channel, and operations, his focus has been consistent: working with the partner community to architect continuity and simplify whitespace deployment from the core data center through to the Edge. Prior experience includes six years of technical sales between earning degrees in mechanical engineering (BSME) and business administration (MBA), both from the University of Notre Dame.

Alex Pope / VP Integrated Rack Solutions - EMEA, Vertiv



The Edge of Mars

What's the most remote IT system in everyday use right now? The Perseverance rover on Mars

Peter Judge
Global Editor

Edge computing is designed to help when applications need a fast response, but are a long way from central IT resources. The most extreme example of this right now is a self-driving vehicle doing detailed science work, 62 million km away from Earth on the surface of Mars.

NASA's Perseverance rover has to handle its environment in real time, but signals take 12 minutes to go from there to NASA's Mission Control. Besides the delay, Internet communications over that distance are unreliable (see Phoning Home), so Perseverance has to be prepared to make a lot of decisions locally.

Despite these demands, the tech deployed to Mars is quite modest: the whole Perseverance rover is managed by the same type of PowerPC 750 processor which powered Apple's Bondi Blue iMac back in 1998.

There's already an installed base on Mars: the Curiosity rover which landed in 2012 has the same processor, and is still in operation. But the Martian environment provides even more compelling reasons to stick with this technology (see box, The Brain)

Due to its more recent design, the tiny Ingenuity drone copter, a passenger on the Mars mission, actually has a somewhat more powerful processor, the Snapdragon 801, which featured in 2014-era smartphones such as the Sony Xperia Z3 (see Flying Buddy).

Yet all this kit is making unbelievable achievements. Even before it begins its scientific study, Perseverance handled its February 18 landing perfectly, analyzing wind patterns and the behavior of its heatshield during its supersonic entry into Mars atmosphere, and then using AI to identify a landing site and steer towards it for touchdown.

The entry, descent and landing (EDL) had to be fully autonomous. The probe plunged through Mars' atmosphere, at a speed of 12,500mph, and a peak temperature of 1,300°C, but NASA engineers on Earth could

Flying buddy

The 1.8kg Ingenuity drone copter, due to make its first flights in April, is the first flying vehicle on Mars, so it's not tasked with any major experiments. It carries a camera for use in scouting the terrain for Perseverance and future rovers, and has a compass, gyro, altimeter and all the sensors needed for autonomous flight - because all Earth-based engineers can do is program in a planned journey.

It's got 4ft diameter rotors - bigger than you'd need on Earth, because the atmosphere is 100 times less dense than ours. It's got six lithium-ion batteries, charged with a solar panel. The copter uses 350W of power, and the batteries store 35Wh, so it's limited to flights a few minutes long, but is expected to travel up to 50m, flying three to five meters from the ground.

A Zigbee radio link gives it 250kbps communication with the rover.

not take a hand at all, because the whole descent took less than seven minutes. Before NASA saw Perseverance start to fall, the rover was already sitting on land.

Entering the atmosphere

NASA has operated five rovers on Mars, but Perseverance was the first to land with its eyes open.

The heatshield and back shell was studded with 28 sensors; for the first four minutes of its descent, the searing temperature and pounding of the atmosphere were recorded by thermocouples, heat flux sensors, and pressure transducers.

When the parachute opened, the heatshield and its sensors was jettisoned. The data was stored for transmission back to NASA - and represents the first detailed data from a Mars landing.

This means future Mars missions can have heatshields designed with data from an actual landing, not a simulation. NASA expects this will allow them to make better heatshields which weigh 35 percent less.

The pressure sensors will tell NASA about the real dynamics of the Martian atmosphere, including the low-altitude winds it hit as it slowed from supersonic speed. Future missions will be able to predict the weather, and landing with more control, in a smaller footprint.

Perseverance's landing target was 4.8 miles by 4.1 miles, already three times smaller than Curiosity's landing target of 15.5 x 12.4 miles. Thanks to the data it captured in February, the next probe will land in a space 30 percent smaller.

Controlled descent

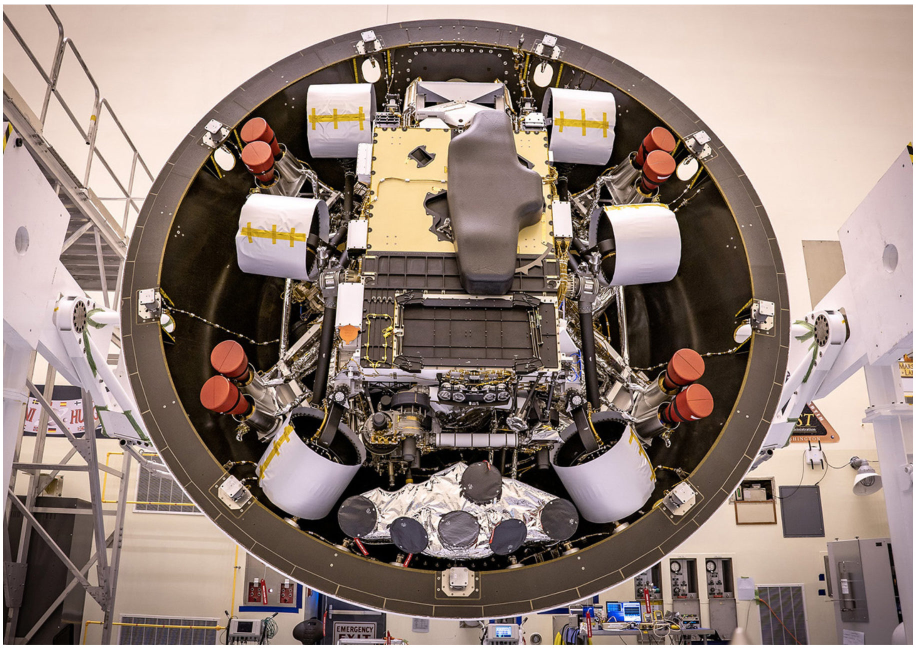
What happened next is even more impressive.

Perseverance is Nasa's fifth rover on Mars - but it's the first to land with its eyes open

As the parachute opened, Perseverance's radar measured its altitude. With the heatshield gone, the rover's cameras could scan the ground. On-board pattern recognition picked out features and looked for the landing spot

When it slowed down to 200mph, the parachute cut loose, and the rover's rockets took over, slowing it right down. At this point the lander vision system (LVS) took over, using "terrain relative navigation" (TRN) to match the rover's camera images to a map of the terrain, and guide it to a smooth landing on the jumbled terrain of Jezero Crater.

The system was tested as much as possible, with helicopters and suborbital rockets on Earth but, for obvious reasons, could not do a full live test till the day of the actual descent.



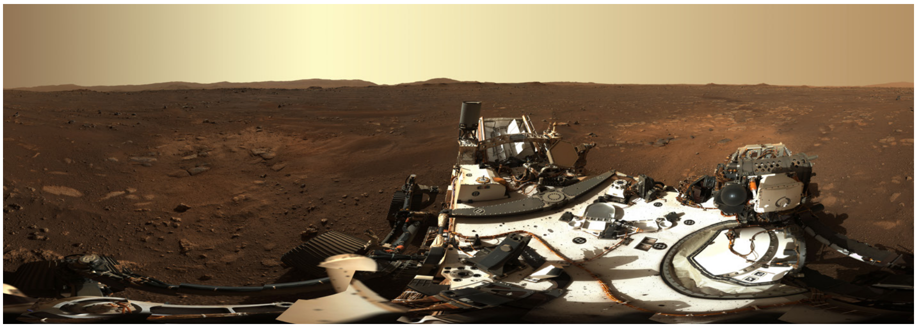
Phoning home

Perseverance communicates back to Earth by relaying signals via Mars orbiters, including the Mars Reconnaisance Orbiter, which has been orbiting Mars since 2006, and became a full time relay system in 2010.

For the leg from Mars orbit to Earth, the interplanetary Internet connection uses a store-and forward-network, designed to deal with frequent errors and disconnects, long delays which can also vary by a huge amount.

Given the light weight of interplanetary craft and their low power budget, communications systems are pretty asymmetrical. Big antennas on Earth are needed to pick up the whispers from Mars, and they must be trained and adjusted to catch signals from pre-arranged directions.

Data is carried by packets defined by the Consultative Committee for Space Data Systems (CCSDS) telemetry standard. Each packet carries a variable amount of data, from 7 to 65,542 bytes, including the packet header. Error correction is also included.



If there ever was life on Mars, Jezero Crater is the best place to look for signs it was once there

Before the landing, NASA's TRN lead Swati Mohan said: "If we didn't have Terrain Relative Navigation, the probability of landing safely at Jezero Crater is about 80 to 85 percent. But with Mars 2020, we can actually bring that probability of success of landing safely at Jezero Crater all the way up to 99 percent every single time."

On the day, when she was the public face of NASA, calling out the telemetry, she said: " it wasn't until after I called 'touchdown confirmed' and people started cheering that I realized, 'oh my gosh, we actually did this. We are actually on Mars. This is not a practice run. This is the real thing'."

Science mission

Jezero is the hardest landing site NASA has chosen for any Mars mission, and it picked it for a reason. Perseverance touched down in an ancient river delta that fed a lake that filled the crater three billion years ago.

If there ever was life on Mars, here is the best place to look for signs of it. Perseverance is kitted out with scientific instruments to look for signs of ancient life in the delta deposits (see box: The toolbag).

It will also drill out and cache interesting rocks for recovery by a later mission. That mission will require whole new techniques, but is due to launch in 2026.

Perseverance will also carry out a key test for possible manned Mars missions in future: testing the production of oxygen from the Martian atmosphere.

All this work will be done more or less autonomously, with high level instructions from Earth bringing back a payload of scientific data.

It really is the farthest Edge computing has ever gone, and embodies several extremes: low data rates, unreliable links, and a "right-sized" processor and memory architecture. It also has absolutely zero chance of any human maintenance and support visits.

Compared to the tight budgets of Perseverance, Earth-bound Edge systems have an embarrassment of riches, with 5G networks, mains electricity and the possibility that someone might come by and reboot them.

While NASA leads scientists round the world in learning from this Mars mission, digital infrastructure builders will be able to learn a lot about the limits of Edge computing.

The Brain

The rover is controlled by a chip that's been in circulation for more than 20 years: the Power PC 750 processor, which also saw service in the vintage Apple iMacs launched in 1998.

It's not a cutting-edge processor: it;s got only 10.4 million transistors, about a thousand times fewer than a smartphone chip. And, while it can run at 233MHz, Perseverance is only operating at 133MHz.

There's a reason for this apparently low spec. It's a ruggedized version of the processor, costing $20,000 and built under license by BAE Systems into the RAD750 single board computer, retooled from the ground up with radiation protection and error correction logic to repair any damage to data in the memory, because a single cosmic ray could fry an unshielded computer.

James LaRosa at BAE Systems told New Scientist: "You have this multi-billion-dollar spacecraft going out to Mars. If it has a hiccup, you're going to lose the mission. A charged particle that's racing through the galaxy can pass through a device and wreak havoc."

The processor runs Wind River's VxWorks real time operating system, which dates back to 1987.

Perseverance has three computers on board, each with two gigabytes of flash memory (about as much as a small USB stick) and 256 megabytes of RAM. One takes care of the rover's main functions, one analyzes navigation images, and the third one is a backup.

The Toolbag

The Mars Rover has the following equipment:

- Mastcam-Z, a panoramic stereoscopic, zooming camera that can help navigation and mineralogy
- SuperCam, an instrument which does chemical analysis and mineralogy at a distance.
- PIXL (Planetary Instrument for X-ray Lithochemistry), an X-ray fluorescence spectrometer which maps the elemental composition of the Martian surface in more detail than ever before.
- SHERLOC (Scanning Habitable Environments with Raman & Luminescence for Organics and Chemicals) the first Raman spectrometer on Mars, it uses an ultraviolet laser to map mineralogy and organic compounds.
- MOXIE (Mars Oxygen In-Situ Resource Utilization Experiment), an experiment to test generating oxygen from the carbon dioxide in Mars' atmosphere. Oxygen will be needed by any future astronauts for breathing - and to burn the rocket fuel which will get them home.
- MEDA (Mars Environmental Dynamics Analyzer), sensors measuring temperature, wind speed and direction, pressure, relative humidity, and dust size and shape.
- RIMFAX (The Radar Imager for Mars' Subsurface Experiment) a ground-penetrating radar to explore the local geology.
- A 2m robot arm, with a rock-coring attachment for gathering samples to be stored in sterile tubes.
- Three antennas operating in UHF (at up to 2Mbps) and X-band.
- A 110W power supply in the form of MMRTG (Multi-Mission Radioisotope Thermoelectric Generator), powered by heat from decaying plutonium-238.

Is OpenRAN in the running?

The Edge will rely on 5G and other radio communications which can be expensive and proprietary. OpenRAN might change that, but it's still being finished

Dan Swinhoe
News Editor

Edge deployments place resources close to applications and the source of their data. But applications like the Internet of Things and autonomous vehicles are so full of moving parts, the only practical way to link them up is through radio networks.

The development of Edge has been closely linked with the arrival of 5G, the short-range high bit-rate evolution of mobile phone networks that is still being delivered. But Edge applications will have to be flexible, and use whatever technology suits their needs - and that could be a problem.

Radio Access Networks (RANs) provide connection between connected devices and the core network via the base stations. Though incredibly important, the technology used is often proprietary so equipment from one vendor will rarely interface with other components from rival vendors. As a result, mobile operators are faced with vendor lock in and use end-to-end solutions from a small set of providers, which can drive up costs and lead to sub-par equipment being used in certain areas.

OpenRAN, however, aims to break down the RAN into component parts and create a unified open interface to connect them. In theory, this allows operators to create bespoke and interoperable best-of-breed deployments. The goal is to create more diversity in the supply chain and allow smaller, specialized companies to enter the market and compete with the incumbents.

Increasing amounts of virtualization, as well as software-defined and cloud architecture in telco infrastructure also means less hardware is required, offering more opportunities for software vendors and more use of commodity off the shelf hardware.

But while there is interest in the technology, is it mature enough for prime time deployment?

Operators like OpenRAN

Overall, everyone *DCD* spoke to said that OpenRAN was generally developing at a decent pace. And while it's been successful in its goal of diversifying the RAN market, challenges remain, especially around interoperability and proven deployments in urban areas or where legacy technology is a consideration.

Dell'Oro Group predicts OpenRAN will account for more than 10 percent of the overall RAN market by 2025 and total $10 billion, but company vice president Stefan Pongratz acknowledges existing suppliers are well positioned to do well with OpenRAN, and the approach won't shift all new investments over to new players.

"OpenRAN is currently trending upwards, although it has yet to reach an inflection point," says Matt Melester, CTO of venue and campus networks at CommScope. "It is succeeding in its broader goals to have the potential of creating a larger ecosystem, but at this point in time, it is too early to tell how successful it will be."

Operators are seemingly keen to at least give OpenRAN an opportunity to mature, as it gives them more leverage over equipment makers. As well as groups like the OpenRAN alliance and the Telecom Infra Project, Deutsche Telekom, Orange, Telefónica, Vodafone, and TIM recently signed a memorandum of understanding around OpenRAN in Europe, signaling their commitment to make it the "technology of choice" for RAN.

Vodafone has been a major supporter of OpenRAN. Last year the company said it planned to deploy OpenRAN technology across 2,600 sites in rural Wales and the South West of England and replace the existing Huawei hardware, with deployment starting in 2022. Andrea Donà, UK network & development director for Vodafone, recently told *Telecom TV* the company had already deployed two OpenRAN sites to its production network as part of its testing process.

Though there are no commercial deployments yet, Vodafone's test and pilot deployments are one of a number currently in development. Outside of the UK, Vodafone is working with Parallel Wireless and others on OpenRAN trials in Turkey, Ireland, and the DRC. Telefonica has an OpenRAN test underway in Peru and Orange in the Central African Republic.

In the US, new wireless mobile provider Dish has plans to cover 70 percent of the US population by June 2023 with its standalone 5G network based on OpenRAN architecture through Fujitsu and Altiostar. In Japan, Rakuten Mobile is rolling out standalone 5G networks using OpenRAN technology in Tokyo, Osaka, and Nagoya. In Germany, Deutsche Telekom is creating an "O-RAN town" in Neubrandenburg and will work with Dell, Fujitsu, NEC, Nokia, Mavenir and others to deploy equipment at 25 "O-RAN compatible sites" that will provide 4G and 5G services.

"All these deployments are using disaggregated network architectures with multiple vendors able to contribute different elements," says John Baker, SVP business development at Mavenir, whose company has been involved in several OpenRAN deployments.

These deployments are important, argues Paul Rhodes, OpenRAN and 5G principal consultant, World Wide Technology (WWT), as they are an opportunity for operators to see and validate good over-the-air performance.

"Rather than theoretically in a lab with a controlled environment, now they're actually exposing it to the real world," he says.

> *"We're having to rip out a lot of Huawei equipment by 2023. We don't have the luxury of waiting for this emergent technology to actually emerge"*

Still work to be done

Mavenir, Parallel Wireless, and Altiostar Networks have found success in the OpenRAN space, while IT infrastructure providers like HPE and Dell are positioning themselves to provide that commodity hardware from which to run virtualized RAN technology. At the same time, incumbent vendors such as Nokia and Ericsson are looking at being involved and are virtualizing some of their offerings.

"Even the traditional network equipment manufacturers, who were governing the space for a long time are now having to open up," says Kalyan Sundhar, VP & GM of 5G edge to core products at Keysight Technologies. "Which tells you that the market is certainly moving in that direction and they have no choice but to move along with it."

However, despite its fans, few believe OpenRAN is ready for prime time deployment in large urban environments yet. Vodafone's Donà acknowledged there was still 'work to be done' around the maturity of the technology, including interoperability, which is a core issue if the multi-vendor 'best of breed' approach is to ever come to fruition. TIM's network engineering director Marco di Costanzo recently told BNamericas it would be "foolhardy" to say OpenRAN is ready for massive roll-out in large centers.

"There are still many hurdles and challenges to overcome, such as supporting advanced features such as carrier aggregation, MIMO, beamforming/steering and others, which require complex, latency sensitive interaction between different RAN blocks," says Prakash Sangam, founder and principal at Tantra Analyst. "OpenRAN has finally graduated from an interesting concept to reality, but it will take considerable time to be mainstream and a default option."

At the same time, purported national security concerns have led some countries, such as the UK, US, and Australia, to exclude Chinese companies such as Huawei and ZTE from new telco network deployments and - in some cases - to rip their equipment out of existing networks.

These moves highlight the need for market diversification, which could benefit OpenRAN - but the timing could be off: the technology may not yet be ready to take full advantage of the switch.

"We're having to rip out a lot of Huawei equipment by 2023 whilst these interface specifications are still getting developed,"says Paul Graham, partner for technology, media and telecommunications at law furn Fieldfisher. "They don't have the luxury of sitting back and waiting for this emergent technology to actually emerge."

Not all the business that was going Huawei is flowing to local incumbents such as Nokia and Ericsson, but the urgency with which some operators need to remove now-forbidden technology means many operators aren't willing to wait for OpenRAN.

"They've got to do it now, and order the equipment now, and it has to be the equipment that's available on the market right now, as opposed to something that might come on them on the market in 12 months' time," he says.

A tale of two OpenRANs

OpenRAN is currently making in-roads on greenfield sites, and that will continue. In theory it can be backward compatible with existing radio networks for 4G, 3G or even 2G, but a lack of mature integration options means standalone 5G OpenRan technology is easier to deploy.

Rakuten in Japan and Dish in the US are opting for greenfield deployments utilizing OpenRAN, and many of the deployments by the incumbent operators are in rural and under-served areas.

Vodafone's first deployment in Wales was at the Builth Wells showground; an area that wouldn't have much capacity requirement for large parts of the year and therefore couldn't previously justify the investment of a large roll-out.

"Early greenfield adopters are more likely to include more components from the broader OpenRAN vision while the migration will be more gradual with the existing networks with initial deployments focusing on the O-RAN interface," says Dell'Oro's Pongratz. "2021 will be a pivotal year for the OpenRAN movement to assess the readiness with brownfield deployments."

"Companies like Rakuten or Dish have taken a different, more proactive approach to OpenRAN," adds Commscope's Melester. "This is because they will not have to start satisfying massive amounts of users right away. They have more latitude to deal with the teething pain of OpenRAN."

Likewise, private LTE deployments could be an area where OpenRAN could find success, partly due to the greenfield nature of such rollouts, and the appeal of removing the need to install fiber or rely on satellite at desired sites. OpenRAN's use of open standards and commodity hardware is also a boon.

"If you're building a network from scratch, and you're not looking for compatibility with anything legacy, then standalone represents a great opportunity [for OpenRAN]," says WWT's Rhodes. "There's a great opportunity for OpenRAN to take an early lead in a sector, and not have to prove itself versus an established competitor."

Mavenir's Baker says his company has been involved with 12 such deployments in 2020, including two 'Industry 4.0' applications in Germany, Naresuan University in Thailand, two indoor pilot projects in Spain, and the Ørsted windfarm in the Irish Sea alongside Vilicom.

"OpenRAN is already well equipped to meet the needs of rural and suburban deployments. Development of some of the more sophisticated technologies required for high demand urban centres is proceeding at pace," he says.

> *"The industry needs to go through some teething pains"*

It's coming, slowly but surely

A number of people *DCD* spoke to predict operators are likely to deploy OpenRAN in greenfield, rural, and standalone networks during 2021 and 2022, and also in private deployments.

"By this time next year, I think everybody will have a pilot deployment that's live and broadcasting over the air," says WWT's Rhodes. "The majority, if not all of the MNOs in any particular country will have OpenRAN presence and will be nodding rather than shaking their head."

Many operability standards are quickly being firmed up – the O-RAN alliance released more than 40 specifications in 2020 – and many of the current stumbling blocks around technology will naturally fall away as the technology matures, and and the first commercial deployments will be rolled out.

"I think we will definitely see some very small targeted deployments [in the next 12 months," says Keysight's Sundhar. "the integration is going to be very daunting, and for it to be a very general-purpose thing is going to take longer."

"The industry needs to go through teething pains. New companies will have more time to work out the bugs, as they don't have legacy infrastructure to support at the same time," says Commscope's Melester, adding that security and power consumption also need work, as well as the interoperability issue.

"In 2021, we'll start to see some of the teething pain associated with real world OpenRAN deployments. This is only natural. 2024-2025 could see parity with traditional legacy OEMs and the gap will start to close in terms of what traditional vendors will be able to produce versus new entrants."

System builders with Edge projects will be watching developments closely, as OpenRAN could be a vital component to turn their ideas into reality.